WASH Data Sharing Update – September 2014

Introduction

Around the world, as many as 1.8 billion people lack access to safe water.[1] This challenge is exacerbated by failures in the water sector resulting in high rates of non-functioning water points. In response to this challenge, governments, NGOs, academics, donors and individuals are increasingly collecting data on the functionality of water points around the world to better understand how to be more impactful. Unfortunately, there is no effective way to share this important data among stakeholders. Costly information gathered is too often used once, by one user, and then remains inaccessible on organizational servers, in dusty reports and in proprietary monitoring systems. Not only is this one-time use of data inefficient in a sector with scarce monitoring resources, but it dramatically limits the opportunity for those addressing water challenges to learn and improve.

WASH Data Sharing Pilot

Given the clear need for improved data communication to enhance decision-making within the water sector, Global Water Challenge (GWC) led the development of a pilot study. The goal of this study was to evaluate the feasibility and value of aggregating disparate data collected through water point mapping initiatives. The study served as a proof of concept, confirming that aggregation of varied datasets under a common platform is possible and analysis of this data will lead to increased understanding.

In conducting the pilot study, GWC developed a draft Data Exchange Standard (DES) based on an analysis of indicators documented in diverse mapping initiatives. A cursory review of publicly-available water point mapping datasets identified 19 indicators that were used in many of the datasets and were included in the draft pilot DES. Four of these ("Update Date", geo-coordinates ("Latitude" and "Longitude"), and "Current Status") were required for data to be included in the pilot.

The pilot study yielded the following learnings:

- The study confirmed that aggregating water point mapping data is feasible. Using the draft DES developed during the study, this exercise brought together nearly 250,000 data points across nine countries and 12 datasets.
- The study confirmed the sector had a need for a platform that enabled connection of water point mapping data. Conversations with varied stakeholders, including national government representatives and NGOs, revealed great support for the initiative and showed different ways the platform could be of value to them. Additionally, the pilot database became one of the

[1] Onda K, LoBuglio J, Bartram J. Global Access to Safe Water: Accounting for Water Quality and the Resulting Impact on MDG Progress. *International Journal of Environmental Research and Public Health*. 2012; 9(3):880-894.

largest publicly-available water point databases and a point of reference for stakeholders such as those in academia.

- The study confirmed the willingness of the sector to work together towards realizing the goals of this initiative. Stakeholders engaged during this study showed readiness to share their data, with some organizations sharing data for inclusion in the pilot database that was previously not publicly-available.

Developing the Data Exchange Standard (DES)

Having confirmed the feasibility, need and support for a water point data sharing platform, GWC expanded its collaboration and began an extended exercise. This exercise aimed to facilitate a discussion within the global water sector and to support the collaborative development of a framework for sharing water point data. This was made possible by support from The Stone Family Foundation. As a starting point for this initiative, a desktop review was carried out to help understand what was currently being mapped and identify common attributes being collected across diverse efforts. Once commonly measured attributes were identified, these could serve as a foundation for a DES.

Call for Data

A request for documented indicators in monitoring exercises was presented to GWC's broad network and the Rural Water Supply Network (RWSN) water point mapping online discussion group (*Dgroup),* among other networks. This call requested that the "header row" of datasets (containing the indicator names) be shared along with a few rows of illustrative data.

Nearly 70 datasets, with a total of over 2,500 indicators, were shared for inclusion in the desktop review. The data was submitted by over 40 unique stakeholders representing implementers, governments and independent consultants.

Data Categorization

Given the magnitude of indicators received for the desktop study, the need for systematic analysis of the data to determine what to include in the baseline DES became apparent. The following categorization process was employed.

Criteria for Inclusion of an Indicator in the Desktop Review

Drinking water system specific: Only indicators providing information specific to a drinking water system were considered in the categorization process.

Objectivity: This was defined as the replicability of the value provided for an indicator by an independent observer. For instance, if an indicator posed a question of "Was the water point crowded?" and the possible answers to the question were "yes" or "no", the indicator was considered subjective and not categorized because another observer could easily find a different value. In contrast, if the indicator provided a value of 50 people/water point, this was considered objective and included in the categorization process.

Dataset uniqueness: Close to 70 datasets were submitted by over 40 individuals, representing a diverse range of organizations. Some organizations provided more than one list of indicators, such as similar surveys used across different country offices. To avoid overrepresentation by one organization in the data, each unique indicator from each organization was included only once in the desktop review.

Structuring Data

Once data was determined to be eligible for inclusion, it was organized in the following structure:

- An "**Area**" was defined as the broad descriptive category within which an indicator would fit.
- An "**Attribute**" was defined as a thematic water point characteristic within a given area.
- An "**Indicator**" was defined as the specific metric measured in a given mapping initiative. This was often the title in the header row in a specific data set.
- A "**Value**" was the information collected from a survey of a specific water point.

Area Identification

The RWSN report, *A Synthesis of Experiences and Lessons discussed in 2012,*[2] formed the initial categorization levels. Areas identified in this report included location details, improvement details, operational details, management details and service levels. The RWSN Dgroup identified categories were further evaluated and revised to five clearly defined areas. Area descriptions are as follows:

- **Location details:** Does this indicator provide information on the location of a water point?
    - E.g. "Village"
- **Infrastructure details***:* Does this indicator provide information on structural or "hardware" components of the water point that are relatively static over time?
    - E.g. "Extraction Technology"
- **Management details***:* Does this indicator provide information on "software" components of a water point that tend to be relatively static over time?
    - E.g. "Installation Year"
- **Water point history***:* Does this indicator provide information on "hardware" or "software" components of a water point that can be dynamic over time?
    - E.g. "Breakdown Year"
- **Collection details***:* Does this indicator provide information on the data collection for a water point?
    - E.g. "Date of Data Collection"

Attribute Identification

After all indicators were grouped into areas, they were grouped with other similar indicators to form an attribute. For example, the attribute "Geo-coordinate –X" refers to the horizontal geographic location of a water point. Indicators such as "Latitude" (decimal), "Latitude" (degrees) and "Northing" were all categorized under this attribute. For each indicator added to the review, if a named attribute existed

[2] Pearce J, C. Howman C. A Synthesis of Experiences and Lessons discussed in 2012. *Rural Water Supply Network (RWSN)*. 2013.

within that area where the indicator fit, the indicator would be grouped in that column. If no attribute existed, that indicator became the basis for a new attribute. As more indicators were added to this new attribute, the attribute name was revised to become more precise. In this way, the list of attributes documented in each area was treated as dynamic during the desktop review. This list was dependent on indicators provided within datasets and could expand as new attributes were identified.

Table 1: Illustrative Data Structure for Desktop Review

Area	**Location Details**		**Infrastructure Details**	
Attribute	Geo-Coordinate - X	Geo-Coordinate - Y	Seasonality	Well Depth
Indicator	Long	Lat	season	totalDepth
Value	0.779	6.010	Dry Always	45

Detailed Value Review

Once all indicators had been grouped within an area and attribute, an analysis was conducted to ensure all values within each attribute answered the same question in the same way. For example, in the case of the attribute "Extraction Technology", each indicator categorized under it must yield values that answer the question of "What system is being used to transport the water from the source to the point of use (e.g. Afridev, gravity scheme, Malda, India Mark II)?" To complete this analysis, all possible values for 12 different attributes were assessed. The attributes reviewed in this detailed review are listed in Appendix B.

This detailed analysis yielded a number of findings. For one, specific attributes were addressed by multiple organizations, but were measured in different ways. As an example, for the attribute "Flow Rate", a number of different units were used for the input values across the surveyed datasets. These ranged from cubic meters per second to number of strokes per 20 liters. On the other hand, indicators that had been initially categorized into different attributes often contained similar values. For example, "Cost Recovery" and "Source of Revenue" attributes often yielded similar values across the surveyed datasets.

Upon reviewing these cases, the attributes were reorganized into different or combined attributes as needed. By looking at how different values were provided for given indicators and attributes, this analysis also helped to identify the advantages and disadvantages of using open text value inputs compared to restricted inputs. This review clarified that while allowing for open text inputs can result in a cumbersome dataset, the specific information held in open text fields can be useful in cases such as describing specific aspects of a water point failure.

This detailed review of the possible values within each attribute provided the basis for how the DES would be structured and what values would be permitted. The specific description for each attribute would aim to ensure that the data collected is sharable but still holds enough specificity to convey meaningful information.

Frequency Tables

With the goal of identifying which attributes were most common, the team determined the number of datasets that provided applicable information for each attribute. This was calculated as

Frequency = (Number of datasets with an indicator corresponding to attribute X) / (Total number of datasets). Thus, an attribute with a frequency of 100% could be addressed with information available in all datasets surveyed.

All attributes currently collected by a majority of frameworks surveyed (i.e. Frequency > 50%) were included in the DES. The reasoning for doing this was to make the DES easily usable given existing practices. A total of 14 attributes were identified to be collected in the majority of frameworks surveyed. These are detailed in Appendix C.

Developing the Initial DES

In addition to the 14 common attributes that the majority of mapping frameworks address, three additional attributes were included in the draft DES to improve the technical functionality of an operational data exchange. These included "Country", "Water Point ID" and "Photo".

These 17 attributes were further clarified to create an operational DES, which would ensure that all attributes could be shared in a consistent and standard way. This included providing guidance to define the attribute, clarifying the acceptable types of values, stating whether data could be transformed if needed (e.g. Easting/Northing to WGS 1984 decimal coordinates) and what values would be acceptable for attributes that would be answered in a list. This additional clarification was based on findings from the detailed value review.

Lastly, three of the selected attributes were identified as required for an operational exchange: "Latitude", "Longitude" and "Date of Data Collection". In addition, for a record to be included, it would need a value for at least one additional attribute from among the remaining 14.

The full DES is available in Appendix A.

Conclusion

The inability of water sector stakeholders, including governments, NGOs, consultants and academia, to share water point data is a critical challenge. However, this challenge is solvable. Among over 25,000 individual indicators from across nearly 70 datasets, much is already common, with 14 attributes being collected by the majority of stakeholders. These attributes, with three additional attributes that ensure operational feasibility of a data exchange, provide a clear and accessible framework for sharing data. Given the commonality of these attributes among existing efforts, exchanging data through this standard is expected to be achievable with minimal effort. This robust foundation will be circulated throughout the water sector for initial input before being finalized. Once finalized, this standard will simply yet effectively allow for unprecedented sharing, learning and improving, fundamentally transforming the way water services are delivered.

Appendices

Appendix A: WASH Data Sharing Draft DES

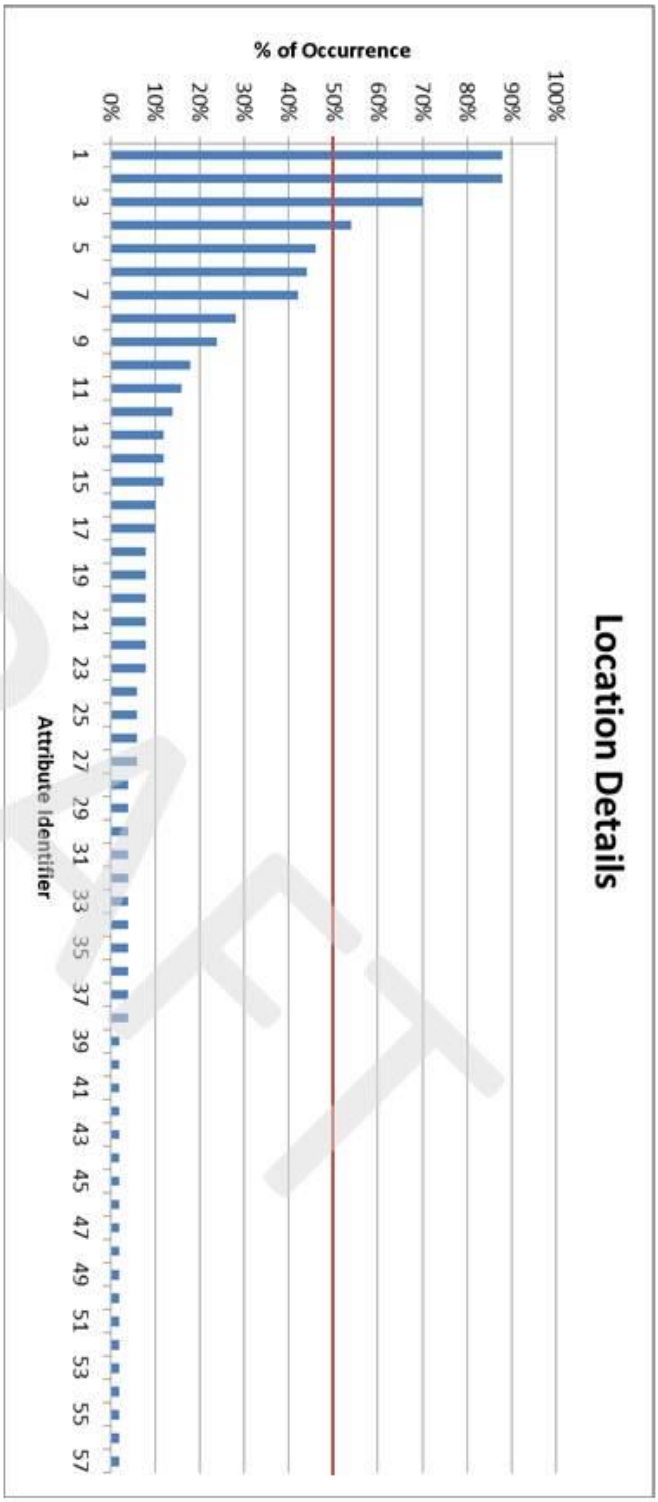
Area	Attribute	Guidance	Values
Location Details	Geo-Coordinate - X	Provide the decimal value of WGS 1984	Numeric
	Geo-Coordinate - Y	Provide the decimal value of WGS 1984	Numeric
	Village	Provide the name of village	Open Text
	District	Provide the name of district	Open Text
	Water Point ID*	Provide the Unique ID for the specific water point infrastructure, as reported by data collector	Open Text
	Country*	Select the ISO two letter country classification code	List (all ISO country codes)
Infrastructure Details	Water Source	Describe the water source (e.g. shallow well, spring, borehole, river, pond, etc.)	Open Text
	Extraction Technology	Describe the approach being used to transport the water from the source to the point of use (e.g. Afridev, gravity scheme, Malda, India Mark II etc.)	Open Text
Management Details	Installation Year	Provide the installation year	Numeric (####)
	Management Structure	Select the entity that manages the water point	List • Direct Government Operation • Private Operator/Delegated Management • Community Management • No Management • Unknown
	Payment Structure+	Select the basis upon which customers pay for water	List • None • Per Container • Per Unit of Time • When Broken • Other
	Cost Per Unit+	Provide the cost per unit as well as the unit with a "/" in between (e.g. "ZAR 10/month", "GHS 15/20L"	Open Text
	Implementer	Provide the name of the entity that installed the water system	Open Text
Water Point History	Presence of Water when Assessed	Identify if any water is available, recognizing that it may be a limited flow	List • Yes • No • Unknown
	Condition	Provide any descriptive status regarding the condition of the water point	Open Text
Collection Details	Date of Data Collection	Provide the date that the data was collected on	Numeric (##/##/####)
Photograph	Photograph*	Provide a photograph of the water system	URL of Photograph

Appendix B: Secondary Analysis Attributes

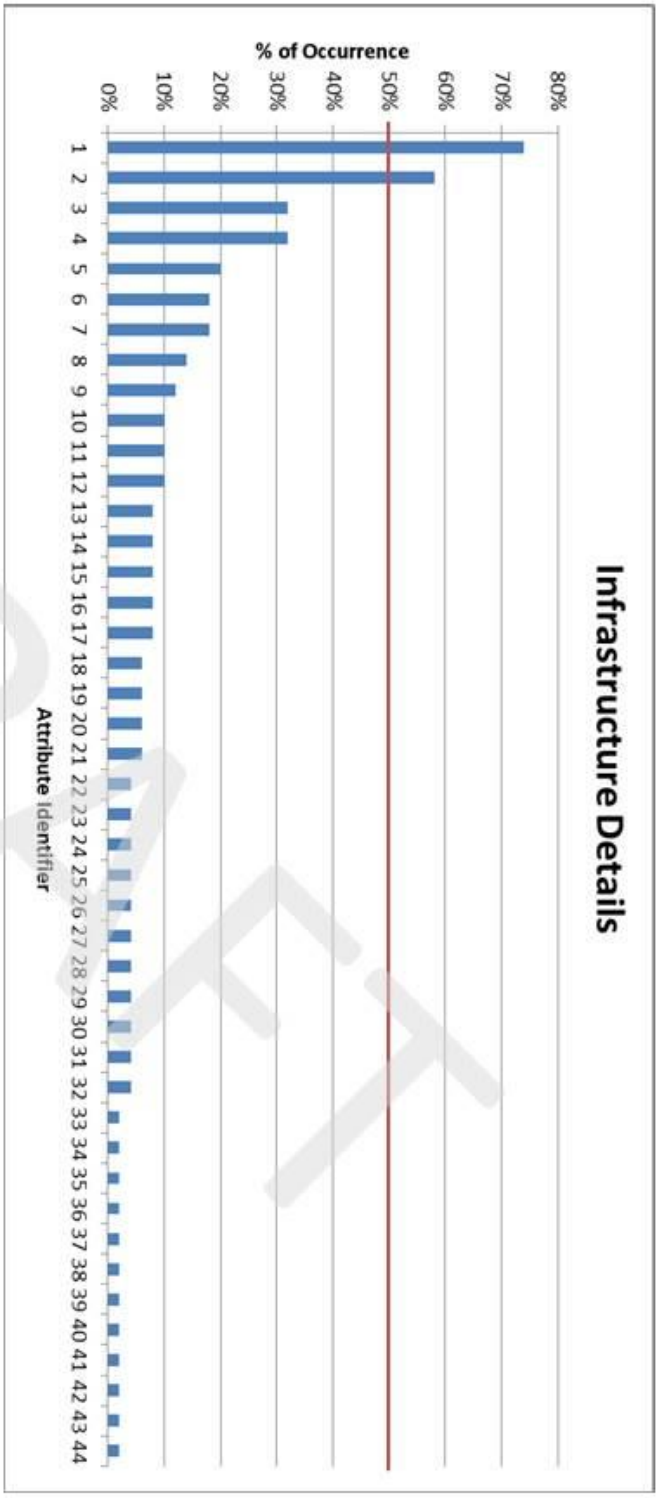
1. Status
2. Functionality
3. Management Unit Active?
4. Management Unit
5. Infrastructure Breakdown
6. Failure Reason
7. Pump Condition
8. Cost Recovery
9. Source of Revenue
10. Breakdown History
11. Breakdown Year/Duration
12. Water Flow Rate

Appendix C: Frequency Analysis

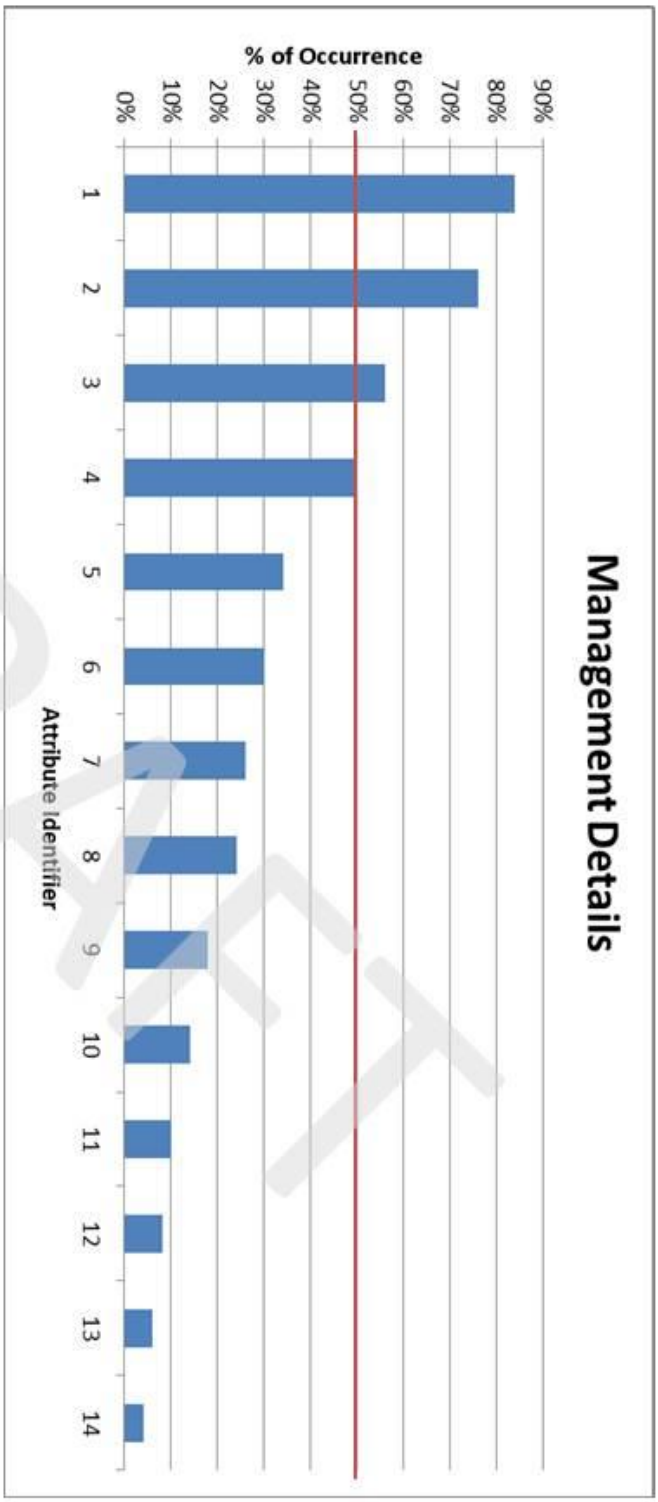
Attribute	Attribute Identifier
Geo-coordinate - X	1
Geo-coordinate - Y	2
Village	3
District	4
Water Point ID (#)	5
Altitude	6
Water point Name	7
Where is the water point located?	8
Region	9
Ward	10
Village ID	11
County	12
Sub-Village	13
Urban	14
Scheme Name	15
Province	16
Rural	17
Country	18
Sub-County	19
sous_prefecture	20
Parish	21
Minicipality	22
Department/State	23
Local Council	24
Section	25
TA	26
Commune	27
Geographic Coordinate System	28
Zone	29
Sub-location	30
prefecture	31
Local Govt. Authority	32
CHIEFDOM	33
Nomadic	34
Division	35
Location	36
Village latitude	37
village longitude	38
Geo Code	39
axe	40
Enumeration Area	41
Settlement	42
Clan	43
Payam	44
Boma	45
Cell	46
Techincal Support Unit	47
Constituency	48
LIA	49
Block	50
Gram Panchayat	51
Zone	52
Woreda	53
Town	54
Sub-city	55
Kebele	56
Upazila Name	57



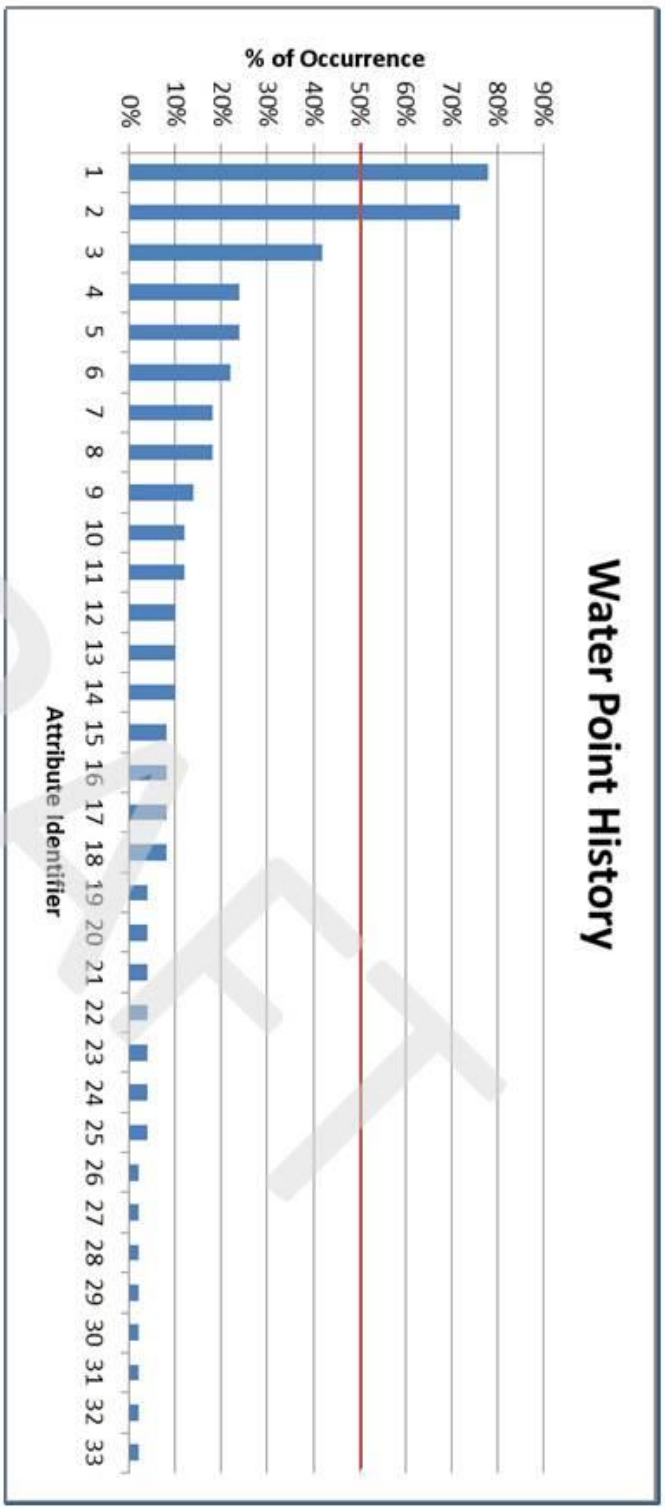
Attributes	Attribute Identifier
Water Source	1
Extraction Technology	2
Water Flow rate	3
Seasonality	4
Water Catchment	5
Water Table Level	6
Well Depth	7
Well Diameter	8
Number of taps per waterpoint	9
Casing Diameter	10
pump depth	11
Tank Capacity	12
Well Casing	13
Casing Length	14
Drainage System	15
Pump Code/Model	16
Connected to a distribution system?	17
Improved/Unimproved Water Source	18
Collection Ponds	19
Fence	20
Tank Condition	21
Water Tube	22
Pump Manufacturer	23
No. of Pumps	24
Drilling Rig	25
Drilling Method	26
Energy Source	27
Metering	28
No. of Tanks	29
Water Strike	30
Household Connections	31
Pipeline Characteristics	32
Casing Stick up	33
Spring Box	34
Well Characteristics - Non specific	35
Platform/Apron Condition	36
Water collection bucket	37
Access Ramps	38
Laundry Block	39
Filter Length	40
Drilling flush Medium	41
Tank Material	42
Bedrock Hit At	43
Distribution points	44



Attribute	Attribute Identifier
Management Unit	1
Installation Year	2
Source of Revenue	3
Implementer	4
Source of Funding	5
Beneficiaries	6
No. of Families	7
Availability of a Mechanic	8
Accessibility by Beneficiaries	9
Local Contact Person	10
Women represenation in Water Committees	11
Location of Spare parts	12
Reliability	13
Financial Records	14



Attribute	Attribute Identifier
Presence of Water When Assessed	1
Condition	2
Breakdown Year	3
Maintenance/Rehabilitation Activity	4
Water Quality - General	5
Turbidity/Color	6
Date of Maintenace/Rehabilitation	7
Fecal Contamination Indicators	8
pH	9
Odor/Smell	10
Source of Pollution	11
Water Treatment	12
electricity conductivity	13
Temperature	14
Name of Rehabilatation Project	15
Total Dissolved Solids	16
Taste	17
What is the waterpoint used for?	18
Cost of Repair	19
Reason for lack of repair	20
Residual Chlorine	21
Iron	22
Fluoride	23
Arsenic	24
Chlorides	25
Alternative sources of Water	26
Silting	27
Leaking	28
Total Hardness	29
Nitrate	30
Surface Film	31
Worms	32
Water Quantity	33



Attribute	Attribute Identifier
Date of Data Collection	1
Comments	2
Data Collector - Person	3
Data Collection ID	4
Data Collector - Organization	5
Verification/Submission Date	6
Interviewee	7
Name of WPM initiative	8
GPS Accuracy	9
GPS Make	10
GPS Model	11
GPS Device	12